JAIST Repository

https://dspace.jaist.ac.jp/

Title	マイクロアレイにより得られる遺伝子発現情報からの 知識発見に関する研究
Author(s)	内藤,隆宏
Citation	
Issue Date	2001-03
Туре	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/740
Rights	
Description	 Supervisor:佐藤賢二,知識科学研究科,修士



Japan Advanced Institute of Science and Technology

Study of knowledge discovery in gene expression data obtained from microarray experiments

Takahiro Naitou

Knowledge Science, Japan Advanced Institute of Science and Technology

March 2001

Keywords: Data mining, Association rule discovery, Microarray, Gene expression.

As the research trend of genome analysis changes from sequencing to gene expression and genetic network identification, microarray has attracted a great deal of attention. However, the methodology for extracting knowledge from a set of microarray data is not established yet. Consequently, the clustering of genes according to their expressivity is still the most popular way of the first analysis to be tried. Since the data obtained from a microarray experiment consists solely of pairs of gene names and their expressivity, it is needed to combine this data with other information extracted from public databases in order to find useful knowledge. To solve this problem, we developed a Web-based analysis system which can discover association rules from a combination of microarray data and public databases[1]. In this study, we describe the functionalities of the prototype system.

The prototype system currently contains 176 sets of data obtained from microarray experiments on *Saccharomyces cerevisiae*, which were provided by Kuhara laboratory at Kyushu University. Each data set consists of approximately 6000 tuples of gene names and expressivity values. In addition, functional information of genes extracted from other two databases (YPD and ENZYME) were integrated with microarray data for finding associations between coexpressing genes and their characteristics. A user can control the protoyupe system by setting various parameters on data preparation and data mining. For example, a threshold value is used to discretize the continuous values of microarray data into three the three states, i.e. inactivated in mutant(-), remains the same, and activated in mutant(+).

^{*}Copyright © 2001 by Takahiro Naitou

Another important functionality of this system is that, before data mining, a user can choose whether or not the transactions are limited to the genes of which the biological function is already known. From the results of some exploratory experiments, we can say that it might be helpful for estimating the function of "functionally unknown" genes. As future works, we are planning to introduce the following improvement to out system.

- Incorporate more microarray data.
- Incorporate more features of genes from other public databases.
- Discard rules of no importance.

References

[1] Takahiro Naitou, Kenji Satou, Emiko Furuichi, Satoru Kuhara and Toshihisa Takagi: A System for Finding Association Rules from Microarray Data and Public Database, Genome Informatics 2000, UNIVERSAL ACADEMY PRESS, INC. TOKYO, JAPAN, pp.356-357.